Natural and Field Experiments: The Role of Qualitative Methods

Thad Dunning
Department of Political Science
Yale University

To be published in *Qualitative Methods*

October 21, 2008

Political scientists increasingly use natural and field experiments in their research.[1] This raises the question—how do qualitative methods contribute to these research methodologies? I suggest here that there are strong complementarities between the use of such research designs and various kinds of qualitative methods. For example, case-based knowledge is often necessary to recognize and validate a potential natural experiment. The research skills associated with qualitative fieldwork, in turn, are often required for the implementation of field experiments. Qualitative methods can be crucial for designing experimental interventions, measuring outcomes, providing evidence on mechanisms, and even constructing random assignment mechanisms.

After discussing natural experiments from a variety of perspectives, I give a short example of how a field experiment may be used to explore the relationship between cross-cutting cleavages and ethnic voting in Mali, drawing on my recent joint research on this topic. As I describe, qualitative methods have contributed in both expected and unexpected ways to this project.

Natural experiments and qualitative methods[2]

An illuminating if well-known exemplar of a successful natural experiment comes from John Snow's studies of cholera transmission (Freedman 1991, 1999, 2005; Dunning 2008). While its substantive domain lies far from the concerns of most social scientists, Snow's research illustrates the key role of qualitative methods in identifying and exploiting a natural experiment to make progress on an important problem.

[1] For evidence on the growing use of field and natural experiments, see Green (2007), Gerber and Green (2008), or Dunning (2008).
[2] Some of the material in this section is based on Dunning (2008); I am grateful to *Political Research Quarterly* and to co-editor Amy Mazur for permission to use the material.

Nineteenth-century London suffered a number of devastating cholera outbreaks. Although predominant theories linked cholera transmission to bad air (miasma) or to ground poisons, Snow became convinced that cholera was a waste- or water-borne infectious disease (Richardson 1887: xxxiv). In Snow's research, “causal process observations” (Collier, Brady, and Seawright 2004) were crucial, both for allowing Snow to formulate a hypothesis about the causes of cholera transmission and to provide evidence for the plausibility of this hypothesis. For example, Snow noted that outbreaks seemed to follow the “great tracks of human intercourse” (Snow 1855: 2); sailors who arrived in a cholera-infested port did not become infected until they disembarked, striking a blow to the miasma theory (Snow 1855: 2).

During London’s cholera outbreak of 1853-1854, Snow famously drew a map showing the addresses of deceased cholera victims. Because these addresses clustered around the Broad Street water pump in the Soho district, Snow argued that contaminated water supply from the pump caused the cholera outbreak. However, there were several anomalous cases: residences located near the pump where there had been no deaths from cholera, and residences far from the pump with cholera deaths. Snow used qualitative process tracing and a heavy dose of “shoe leather” (Freedman 1991) to probe these seemingly disconfirming outcomes (Snow 1855: 39-45). At a brewery located near the Broad Street pump, where cholera death rates were anomalously low, the proprietor told Snow that a fresh-water pump was installed on the premises—and that in any case the brewers tended to drink beer, not water (Snow 1855: 42). At another address, closer to another water pump than to Broad Street—and where there had been significant deaths from cholera—Snow learned that the deceased residents had preferred, for one reason or another, to take water at the Broad Street pump (Dunning 2008). Snow’s experience as a clinician, his studies of the pathology of cholera deaths, and his spot map showing the proximity of victims to the Broad Street pump all provided bits of evidence, which suggested that cholera might indeed be an infectious disease carried by waste or water.

However, Snow's most powerful piece of evidence came from a natural experiment. Large areas of London were served by two water suppliers, the Lambeth company and the Southwark and Vauxhall company. Just prior to the cholera epidemic of 1853-54, the Lambeth company moved its intake pipe further upstream on the Thames, thereby "obtaining a supply of water quite free from the sewage of London" (Snow 1855: 68), while the Southwark and Vauxhall company left its intake pipe in place. After painstaking data collection, Snow constructed a simple cross-tab showing cholera death rates during the epidemic by source of water supply. For houses served by Southwark and Vauxhall, the death rate from cholera was 315 per 10,000; for houses served by Lambeth, it was a mere 37 per 10,000 (Snow 1855, Table IX, p. 86; presented in Freedman 2005).

Why did this constitute a credible natural experiment? Unlike true experiments, the data used in natural experiments come from naturally occurring phenomena – actually, in the social sciences, from phenomena that are often the product of social and political forces. Because the manipulation of the treatment, intervention, or independent variable is not generally under the control of the analyst, natural experiments are, in fact, observational studies. However, unlike other non-experimental approaches, a researcher exploiting a natural experiment can make a credible claim that the assignment of the non-experimental subjects to treatment and control conditions is "as-if" random. Outcomes are compared across treatment and control groups, and both *a priori* reasoning and empirical evidence are used to validate the assertion of randomization.

Thus, random or as-if random of assignment to treatment and control conditions—in Snow's study, the water supply source—constitutes the defining feature of a natural experiment. This implies that at least as a necessary if not sufficient condition, the treatment and control groups are balanced with respect to other (measurable) variables that might explain cholera deaths. Notice that in a natural experiment, this is achieved not by statistical adjustment on the part of the analyst but rather by nature's as-if randomization. Snow presented various sorts of evidence to establish this pre-treatment equivalence between the two groups. In his own words,

> The mixing of the (water) supply is of the most intimate kind. The pipes of each Company go down all the streets, and into nearly all the courts and alleys. A few houses are supplied by one Company and a few by the other, according to the decision of the owner or occupier at that time when the Water Companies were in active competition. In many cases a single house has a supply different from that on either side. Each company supplies both rich and poor, both large houses and small; there is no difference either in the condition or occupation of the persons receiving the water of the different Companies…It is obvious that no experiment

> could have been devised which would more thoroughly test the effect of water supply on the progress of cholera than this" (Snow 1855: 74-75).

Moreover, residents did not appear to self-select into their source of water supply: decisions regarding water companies were often taken by absentee landlords, the decision of the Lambeth company to move its intake pipe was taken before the cholera outbreak of 1853-54, and existing scientific knowledge did not clearly link water source to cholera risk. As Snow puts it, the pipe's move meant that more than three hundred thousand people were:

> divided into two groups *without their choice, and, in most cases, without their knowledge*; one group being supplied with water containing the sewage of London, and…the other group having water quite free from such impurity (Snow 1855: 75; italics added).

The cholera example provides several useful lessons about the elements of a successful natural experiment (see Freedman 1991, 1999). Snow went to great lengths to gather evidence and to use *a priori* reasoning to argue that only the water supply distinguished houses in the treatment group from those in the control group, and thus the impressive difference in death rates from cholera was due to the effect of the water supply. It is also worth noting that, while the natural experiment may have been the *coup de grace* in Snow's painstaking investigation into the causes of cholera transmission, his use of this natural experiment was complemented and indeed motivated by the other evidence that he had gathered. The body of evidence Snow compiled depended on his detailed knowledge of the progress of previous cholera outbreaks in England, on his ability to cull information from a variety of sources, and especially on his willingness to do on-the-ground process tracing and close-range exploration of seemingly disconfirming cases (Dunning 2008). This kind of close-range research also gave him the information he needed to discover and exploit his natural experiment, while his apparently innate sense of good research design led him to recognize the inferential power of the approach.

Social-scientific examples

Several of the elements of Snow's successful natural experiment can be found in recent social-science applications as well. Brady and McNulty (2004), for example, are interested in examining how the cost of voting affects turnout. In California's special gubernatorial recall election of 2003, in which Arnold Schwarzeneggar became governor, the elections supervisor in Los Angeles County consolidated the number of district voting precincts from 5,231 (in the 2002 regular gubernatorial election) to 1,885. For many voters, the physical distance from residence to polling place was increased, relative to the 2002 election; for others, it remained the same. Those voters whose distance to the voting booth changed – and who therefore presumably had higher costs of voting, relative to the 2002 election -- constituted the treatment group, while the control group voted at the same polling place in both elections.

The consolidation of polling places in the 2003 election arguably provides a natural experiment for studying how the costs of voting affect turnout. A well-defined intervention, the closing of some polling places and not others, allows for a comparison of average turnout across treatment and control groups. The key question, of course, is whether assignment of voters to polling places in the 2003 election was as-if random with respect to other characteristics that affect their disposition to vote. In particular, did the county elections supervisor close some polling places and not others in ways that were correlated with potential turnout? Brady and McNulty (2004) raise the possibility that the answer to this question is yes, and indeed they find some evidence for a small lack of pre-treatment equivalence on observed covariates such as age across groups of voters who had their polling place changed (i.e., the treatment group) and those that did not. Thus, the assumption of as-if random assignment may not completely stand up either to Brady and McNulty's careful data analysis or to *a priori* reasoning (elections supervisors, after all, may try to maximize turnout). Yet pre-treatment differences between the treatment and control groups are small, relative to the reduction in turnout associated with increased voting costs. After careful consideration of potential confounders, Brady and McNulty can convincingly argue that the costs of voting negatively influenced turnout, and a natural experimental approach plays a key role in their study.

Another increasingly common class of natural experiments exploits the existence of political or jurisdictional borders that separate similar populations of individuals, communities, firms, or other units of analysis, some exposed to a treatment or policy intervention and others not; in Dunning (2008), I review several studies and discuss the strengths and limitations of this form of natural experiments. Posner (2004), for example, studies the question of why cultural differences between the Chewa and Tumbuka ethnic groups are politically salient in Malawi but not in Zambia. Separated by an administrative boundary originally drawn by Cecil Rhodes' British South African Company and later reinforced by British colonialism, the Chewas and the Tumbukas on the Zambian side of the border are apparently identical to their counterparts in Malawi, in terms of allegedly objective cultural differences such as language, appearance, and so on. However, Posner finds very different inter-group attitudes in the two countries, with Chewas and Tumbukas in Malawi more likely to report an aversion to inter-group marriage and a disinclination to vote for members of the other group.

Posner argues convincingly that long-standing differences between Chewas and Tumbukas located on either side of the border cannot explain the very different inter-group relations in Malawi and in Zambia; a key claim is that "like many African borders, the one that separates Zambia and Malawi was drawn purely for [colonial] administrative purposes, with no attention to the distribution of groups on the ground" (Posner 2004: 530). Instead, the factors that make the cultural cleavage between Chewas and Tumbukas politically salient in Malawi but not in Zambia should presumably have something to do with exposure to a treatment (broadly conceived) on one side of the border but not on the other. Posner suggests that contrasts between inter-group attitudes of Chewas and Tumbukas in Malawi and Zambia are explained by the different sizes of these groups in each country, relative to the size of the national polities, which changes the dynamics of electoral competition and makes the groups political allies in Zambia but rivals in Malawi (see also Posner 2005).

Yet in order to argue this, Posner has to confront a key question which, in fact, sometimes confronts randomized controlled experiments as well: what, exactly, is the treatment? Or, put another way, which aspect of being in Zambia as opposed to Malawi causes the difference in political and cultural attitudes? Posner provides evidence that helps rule out the influence of electoral rules and the differential impact of missionaries on each side of the border. Rather, he suggests that in Zambia, Chewas and Tumbukus are politically mobilized as part of a coalition of Zambians living in the country's Eastern region, since alone neither group has the size to contribute a substantial support base in national elections, whereas in smaller Malawi (where each group makes up a much larger proportion of the population), Chewas are mobilized as Chewas and Tumbukus as Tumbukus (see also Posner 2005).

Clearly, the hypothesized intervention here is on a large scale – the counterfactual would involve, say, changing the size of Zambia while holding constant other factors that might affect the degree of animosity between Chewas and Tumbukus. This is quite different from imagining changing the company from whom one gets water in nineteenth-century London; one may question whether a manipulationist account of causation is most appropriate here (see Goldthorpe 2001 and Brady 2002). However, Posner's investigation of the plausibility of the relevant counterfactuals provides an example of "shoe leather" (that is, walking from house to house to find nuggets of evidence and rule out alternative explanations) in the tradition of John Snow (Freedman 1991).

In natural experiments, a key question is whether treatment assignment really is as-if random, that is, independent of other factors that might explain differences in average outcomes across treatment and control groups. The assertion of as-if random assignment may be more compelling in some contexts than in others. As I discuss in Dunning (2008), it may be useful to conceptualize a "continuum of plausibility" that assignment to treatment and control is really as-if random; in that article, I place several recent studies along such a continuum and discuss ways in which the as-if random criterion may be partially validated with evidence as well as *a priori* reasoning (Dunning 2008).

For present purposes, the central point is simply that qualitative methods and case-based knowledge may play an important role in efforts to exploit as well as to validate natural experiments. Close knowledge of specific substantive domains may allow analysts to find and exploit credible natural experiments (see also Malesky, this symposium). And while simple quantitative techniques are also important for partially validating the claim of as-if random assignment (for example, for demonstrating equivalence on measured non-treatment variables across treatment and control groups), leveraging case-based knowledge about the substantive domain under investigation is also crucial to convincing applications of the natural-experimental approach.

Field experiments and qualitative methods

In a randomized controlled experiment, subjects or units are randomized to treatment and control, and the intervention or manipulation is under the control of an experimental researcher (Freedman, Pisani, and Purves 1997). The main attraction of true (randomized controlled) experiments is that they solve pervasive problems of confounding and selection bias: random assignment ensures that treated and untreated groups are equivalent prior to the intervention, up to random error.[3] With a large enough number of units, random error will play only a small role, and post-intervention differences across the treatment and control groups can be reliably attributed to the effect of treatment.

Field experiments—that is, randomized controlled experiments in which the "conditions under which a causal process of interest occurs are simulated as closely as possible" (Gerber and Green 2008)—offer many synergies with qualitative methods. As Gerber and Green (2008) point out, by definition, field experiments constitute "the conjunction of two methodological strategies, experimentation and field work." In some obvious ways, then, the skills associated with some qualitative researchers, particularly those who do fieldwork, are requisite for field experiments as well. The close case-based knowledge associated with some qualitative research may be vital for recognizing the opportunity to conduct a field experiment, and the social and networking skills often associated with qualitative fieldwork appear to be the *sine qua non* of many field experiments as well.

Qualitative methods may play several other important roles in field experiments, however. Although not my main focus here, one important potential contribution of qualitative methods is in identifying mechanisms, which is a crucial part of causal inference. For example, an experiment may allow the estimation of a causal effect without, however, illuminating the mechanism through which the cause produces its effect. Qualitative information may provide insights or information on context and mechanism, perhaps in the form of what Collier, Brady, and Seawright (2004) call

[3] Of course, problems of post-intervention bias can arise: subjects who get the vaccine may tend to go swimming.

"causal process observations." (In addition, other experiments might be designed to elucidate the mechanism).

Yet there are also many other ways in which qualitative methods can contribute to field experiments, beyond simply field research skills. For example, they can help analysts confront challenges involved in measuring outcomes, designing treatments, recruiting participants, and even randomizing subjects to treatments. My objective in the rest of this article is to describe the contributions of qualitative methods to an ongoing experiment on ethnic politics in Mali. I first describe the experiment briefly, in order to set the stage for my discussion of qualitative methods.

Cross-Cutting Cleavages and Ethnic Politics: An experiment in Mali

Social scientists often ascribe the absence or moderation of ethnic conflict to cross-cutting cleavages—that is, the presence of alternate dimensions of identity or interest, along which members of the same ethnic group may have diverse allegiances. Despite a rich theoretical literature, however, the empirical effects of cross-cutting cleavages are notoriously difficult to estimate. One goal of my ongoing research, conducted jointly with Yale undergraduate Lauren Harrison, is to formulate an experimental method for investigating the political effects of cross-cutting cleavages.

In Mali, despite substantial ethnic diversity, levels of ethnic conflict are persistently low. Unlike some Sub-Saharan countries, parties do not form along ethnic lines, and ethnicity is a poor predictor of individual vote choice. One set of explanations advanced for this African anomaly focuses on an informal institution called *cousinage* (loosely translated as "joking cousinship"). In Mali as well as in Sénégal, the Gambia, Guinea, western Burkina Faso, and the northern Ivory Coast—areas either formerly part of the Mali Empire (c. 1230-1600) or subject since to significant immigration from those areas—families historically formed alliances on the basis of patronyms. These historical alliances are now invoked in everyday social interactions. Today in Mali, for instance, if someone with the last name Keita meets someone named Coulibaly on the street, these two fictive cousins may invoke a standard set of jokes, even if they have never previously met. The jokes reinforce the social bonds understood to inhere in their relationship.

For our purposes, these alliances constitute cross-cutting cleavages, because they occur across as well as within ethnic groups.[4] Despite a substantial literature on the alleged pacifying effects of cousinage (see Canut and Smith 2006; Davidheiser 2006: 837; Launay 2006; among early anthropologists, Mauss 1928 and Radcliffe-Brown 1940), it appears to us that this claim has been not been subjected to empirical scrutiny that would allow valid inferences about causal effects. We extend the hypothesis to explain not only the absence of ethnic conflict, generically, but also the apparent absence of ethnicity in electoral politics, asking why, in an ethnically-diverse African polity, ethnicity not predict individual vote choice, and parties do not form along ethnic lines. Our extension of the cross-cutting cleavage (cousinage) hypothesis to explain political preferences and patterns of electoral competition in Mali is new and to our knowledge has not been previously tested.

We developed an experimental design to estimate the effects of cousinage relations on evaluations of political candidates and their speeches. First, we videotaped two Malian actors delivering the same speech, which focused on standard themes in Malian political campaigns; in initial field trials in the capital of Bamako, 56% percent of experimental subjects said the speech "reminded them of a speech they had heard on a previous occasion." The speech was delivered in Bambara, which is the lingua franca of Bamako (and of Mali).[5] We then recruited experimental subjects by canvassing all of Bamako's neighborhoods (*quartiers*), approaching men and women sitting outside homes (or knocking on doors) and asking subjects if they would participate in a study on political speeches.[6] We administered a screening questionnaire to each potential subject, asking for each subjects' first and last name and ethnic identity, along with various other personal information; this allowed us to assign subjects randomly to the treatment conditions, as described below.[7] Experimental subjects then viewed our videotaped

[4] For example, the Keita are part of the Malinké ethnic group, while their joking cousins the Coulibaly are part of the Bambara ethnic group.
[5] Though Bambara is the first language of one ethnic group in Mali, its use does not imply a particular ethnic identity on the part of the politician. When experimental subjects were not provided with the politician's last name, their guesses about his ethnicity closely tracked the distribution of ethnic groups in Bamako.
[6] The experimental population is a convenience sample, but distributions on several measured variables are similar to those given by the census for Bamako. However, the experiment under-represents women.
[7] First name and other identifying information of subjects was subsequently discarded, as described in our protocol approved by Yale's human subjects review board.

political speeches on a portable DVD player or laptop, using headphones.[8] Finally, subjects then answered questions about the content of the speech and the politician who delivered it. For instance, they answered questions about the global quality of the speech, whether the speech made them want to vote for the candidate, and specific questions about candidate attributes such as competence, likeability, and intelligence.

The manipulation in this experiment consisted of what subjects were told about the politician's last name. In Mali, last name conveys information about both ethnic identity and about cousinage ties. Thus, varying the politician's last name allowed us to vary the treatment along two dimensions: the ethnic relationship of the politician and the subject (same ethnicity/different ethnicity) and their cousinage relationship (joking cousins/not joking cousins). Our resulting experimental design had six treatment conditions, four of which are shown in the cells of Table One. We also added a fifth condition, in which the subject was provided with no information about the last name of the politician (and thus no information about ethnicity or cousinage ties), and a sixth treatment condition, in which the politician had the same last name as the subject.[9]

Table 1: Experimental Design (Four of Six Treatments)

	Joking cousins	Not joking cousins
Same ethnicity		
Different ethnicity		

According to our hypotheses, a joking cousin relationship between voters and politicians should moderate the negative effect of ethnicity on voters' evaluations of politicians. We expect evaluations of politicians to be more positive on average if the

[8] Only experimental subjects could hear the speech through the headphones, and only one subject was recruited from any group; subjects also answered follow-up questions on their own. This limited the potential that subjects' responses to treatment depended on the treatment assignment of other subjects.
[9] The sixth treatment may allow us to distinguish a "same ethnicity" or a "joking cousin" effect from a mere "sameness" effect: perhaps people simply want to vote for politicians who share their last names.

politician is a co-ethnic: thus, in Table 1, we expect to find that mean evaluations of co-ethnic politicians (first row) are more positive than mean evaluations of non co-ethnics (second row). On the other hand, we also expect joking cousins to be evaluated more positively than non-joking cousins, so that mean evaluations of subjects in the first column are more positive than evaluations in the second column. The main point, however, is that we expect non-coethnic cousins (top-right cell) to be evaluated more positively than non-coethnic, non-cousins (bottom-right cell).[10]   Such a finding would be consistent with the idea that due to cousinage relations, members of the same ethnic group have diverse allegiances along a cross-cutting dimension of identity.[11]

We began rolling out this experiment at the end of July 2008; though we have finished initial field-testing at the time of writing, we have not yet seen data from the main phase of data collection. The publication of hypotheses in this newsletter constitutes a public posting of the experimental protocol prior to analysis of the data. Our principal form of analysis for testing these hypotheses will be difference-of-means tests across subjects randomly assigned to each of the six treatment conditions, with ancillary testing of sub-groups due to our interest in possible treatment effect heterogeneity.

In the interest of brevity, I will now describe just two areas in which qualitative methods have been crucial in designing and implementing this experiment: the design of the experimental stimulus, and the creation of a cousinage matrix that allowed us to assign subjects to treatment conditions.

The experimental stimulus: writing a typical political speech

Our goal in designing the experimental stimulus was to create a speech that would engage subjects' attention while mimicking as closely as possible a typical political speech given by a candidate for deputy in the legislature. Here, one of us (Lauren Harrison) drew on earlier fieldwork in which she observed parliamentary campaigns in Bamako in 2007. After comparing our speech to transcripts of real political speeches, we vetted the speech with several Malian informants. I will not belabor the point here but will simply point out that fieldwork and other qualitative methods played an important

[10] However, based on our qualitative research, we believe that subjects may not clearly distinguish between cousins and non-cousins, among their co-ethnics.
[11] We do not have strong expectations about the sign of any interaction between co-ethnicity and cousinage.

role in the design of the experimental treatment.

Random assignment: creating a cousinage matrix

More involved fieldwork was required for the second topic I will discuss here. In order to assign subjects at random to one of the six treatment conditions, we created a large matrix, each row of which corresponds to a Malian last name that we could expect to encounter in the field.

For instance, Table Two shows a row of the matrix for a person named Keita from the Malinké/Maninka ethnic group. The columns of this row give the last names associated with each of our six treatment conditions. For example, the names in the first two columns are all from the same ethnic group, but Sissoko and Konaté (first column) are considered cousins of the Keita, while Diané (second column) is not. The names in the third and fourth columns, on the other hand, are names associated with other ethnic groups, some of them cousins of the Keita (third column) and some of them not (fourth column). Note that in cells with multiple entries, such as in the first, third, and fourth column in Table Two, the politician's assigned last name was selected at random from the names in the cell.

Table Two: A typical row of our random assignment matrix

	(1) Co-ethnic/ Cousin	**(2) Co-ethnic/ Not cousin**	**(3) Not co-ethnic/ Cousin**	**(4) Not co-ethnic/ Not cousin**	**(5) No name**	**(6) Same name**
Keita (Maninka)	1. Sissoko 2. Konaté	1. Diané	1. Doucouré 2. Sacko 3. Sylla 4. Coulibaly 5. Touré	1. Diallo 2. Cissé 3. Dambelé 4. Théra 5. Dabo 6. Togola 7. Watarra	Pas de nom	**Keita** (Maninka)

Qualitative fieldwork was crucial for constructing this cousinage matrix. Before arriving in Bamako, we reviewed the secondary literature and conducted interviews with experts on cousinage as well as ordinary Malian informants. This enabled us to determine, as an initial matter, the cousins that are associated with many Malian last names and to construct a preliminary, skeletal matrix. Upon arrival in Mali, we solicited feedback on the matrix from key informants and, with their help, added to the list of names included in the left column (that is, the names of potential subjects) and also refined the list of politicians' names included in each column of each row.

Next, we field-tested an initial version of the matrix on 169 subjects. Data from this initial field trial, as well as additional qualitative information obtained in the field, allowed us to expand and improve the matrix again, and 47 more subjects participated in a second phase of the experiment using our improved matrix. Finally, in mid-August 2008, we revised the matrix once again, for reasons discussed below; this final revised matrix is being used to roll out the experiment during September 2008. Our final version of the matrix includes more than 200 names in the left-hand column, including all of the most typical Malian names.

In our initial field trials, experimental subjects did not always perceive themselves to be in the correct cell—that is, the treatment condition to which they had been randomly assigned. In fact, subjects inferred ethnicity with great accuracy: given only the last name of the politician, and choosing from more than 14 possible ethnic categories, subjects correctly classified the politician's ethnicity 75% of the time. However, in initial trials, they more frequently labeled cousins as non-cousins, or non-cousins as cousins.

This mismatch in initial trials between the treatment conditions to which some subjects were assigned, and the treatment conditions they perceived, raises important inferential issues.[12] After all, what we care about in this study is the effect of subject perceptions -- we want to know how *perceiving* oneself as being a cousin or not being a cousin of the politician, or his co-ethnic or not, shapes evaluations of the candidate's speech. Here, the mismatch probably occurred for two reasons. First, correctly classifying cousinage relations for over 200 last names is difficult; our initial matrix of cousinage relations was highly imperfect. In this experiment, there was a tradeoff involved in limiting the names of potential subjects. On the one hand, cousinage

[12] From an experimental design perspective, this issue can be analogized to the problem of compliance with an experimental protocol. See Freedman (2006) for a discussion of relevant analytic approaches.

relations are much better understood by us (and by Malians) for a few very common names, such as Keita, Coulibaly, Touré, or Cissé, than for less common names, so we might have had a better overall accuracy/compliance rate had we limited the study population to subjects with such last names. On the other hand, limiting the number of names would have meant more inefficient and costly subject recruitment.

Second, however, even if we could create a perfectly accurate matrix of cousinage relations, as understood by key informants, people vary in their knowledge of cousinage relations in Mali. For instance, are the Keita and the Doucouré (third column of Table Two) really cousins? Reasonable minds can apparently disagree. As one leading expert on cousinage puts it, "The question of which *jamu* [patronym] actually jokes with whom is subject to considerable indeterminacy. Lists of the joking partners of any given *jamu* may vary from community to community, or even from individual speaker to speaker" (Launay 2006: 799). Our own experience in the field validated this observation.

The key to resolving this conundrum is that some cousinage links are in fact widely understood: everyone agrees that the Keita and the Coulibaly are cousins. We therefore took the approach of limiting names in the first and third column of Table Two to those *vrai cousins* or true *senanku* (the Bambara word for cousin), while also only including names in the second and fourth cell that we thought would maximize the chance of correct identification as non-cousins. We devoted considerable effort in the field to accomplishing this task, with the help of key informants. Initial indications suggest that our revised cousinage matrix is allowing much greater accuracy in subject assignment to treatment during the main roll-out of the experiment.

The point is that eliciting a reliable map of cousinage relations from key informants very centrally involved qualitative as well as mixed methods. For instance, to revise our cousinage matrix we conducted qualitative interviews with key informants. We then also employed quantitative analysis of the experimental data from initial trials. To improve the cousinage matrix, we therefore iterated between focused interviews, new versions of the cousinage matrix, and our experimental data to improve the random assignment mechanism in this experiment.

Finally, qualitative methods will likely play a key role in interpreting the results of the experiment—for example, in assessing the extent to which the experimental results can allow us to infer that cousinage plays the political role attributed to it. Here, we will want to analyze the potential role of cousinage in important parliamentary and presidential electoral campaigns.

Conclusion

Natural and field experiments are assuming a place of greater prominence in political science. They also appear to offer substantial opportunities to qualitative researchers. The type of experiment I described in Mali can be implemented relatively inexpensively; in fact, such a project would probably be well within reach for a graduate student working on his or her dissertation. Most importantly for present purposes, natural and field experiments often require skills and case-base knowledge associated with qualitative research. The inferential advantages of natural and field experiments may be increasingly combined with the strengths of qualitative research to generate new forms of mixed-method research, in the service of research programs in many different substantive areas.

References

Arceneaux, Kevin, Donald Green and Alan Gerber. 2006. "Comparing Experimental and Matching Methods Using a Large-Scale Voter Mobilization Experiment." *Political Analysis* 14: 37-62.

Berk, Richard, and David Freedman. 2008. "On weighting regressions by propensity scores." *Evaluation Review* 32: 392-409. Available online at http://www.stat.berkeley.edu/~census/weight.pdf

Brady, Henry E. 2002. "Models for Causal Inference: Going Beyond the Neyman-Rubin Holland Theory. " Paper presented at the Annual Meeting of theAPSA Political Methodology Working Group, Seattle, Washington, 16 July.

Brady, Henry E. and John McNulty. 2004. "The Costs of Voting: Evidence from a Natural Experiment." Paper presented at the annual meeting of the Society for Political Methodology, Stanford University, July 29-31, 2004.

Canut, Cécile, and Étienne Smith. 2006. "Pactes, Alliances et plaisanteries. Pratiques locales, discours global." In Cécile Canut and Étienne Smith, eds. "Parentés, plaisanteries et politique," special issue of *Cahiers D'Études Africaines* XLVI (4): 795-808.

Collier, David, Henry E. Brady, and Jason Seawright. 2004. "Sources of Leverage in Causal Inference: Toward an Alternative View of Methodology." Chapter 13 in *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Rowman & Littlefield.

Davidheiser, Mark. 2006. "Joking for Peace: Social Organization, Tradition, and Change in Gambian Conflict Management." In Cécile Canut and Étienne Smith, eds. "Parentés, plaisanteries et politique," special issue of *Cahiers D'Études Africaines* XLVI (4): 795-808.

Dehejia, Rajeev. 2005. "Practical Propensity Score Matching: a reply to Smith and Todd." *Journal of Econometrics* 125 (1): 355-364.

Dehejia, Rajeev H., and Sadek Wahba. 1999. "Causal Effects in Nonexperimental

Studies: Reevaluating the Evaluation of Training Programs." *Journal of the American Statistical Association* 94: 1053-1062.

Doherty, Daniel, Donald Green, and Alan Gerber. 2006. Personal Income and Attitudes toward Redistribution: A Study of Lottery Winners. *Political Psychology* 27 (3).

Dunning, Thad. 2008. "Improving Causal Inference: Strengths and Limitations of Natural Experiments." *Political Research Quarterly* 61 (2): 282-93.

Dunning, Thad, and Susan Hyde. 2008. "The Analysis of Experimental Data: Comparing Techniques." Paper presented at the annual meetings of the American Political Science Association, Boston, MA, August 27-September 2, 2008.

Freedman, David. 1991. "Statistical Models and Shoe Leather." In P.V. Marsden, ed., *Sociological Methodology,* Vol. 21. Washington, D.C.: The American Sociological Association.

Freedman, David. 1999. "From association to causation: Some remarks on the history of statistics." *Statistical Science* 14: 243-58.

Freedman, David. 2005. *Statistical Models: Theory and practice.* Cambridge: Cambridge University Press.

Freedman, David. 2006. "Statistical Models for Causation: What Inferential Leverage Do They Provide?" *Evaluation Review* 30: 691-713.

Freedman, David, Robert Pisani, and Roger Purves. 1997. *Statistics.* 3rd 3d. New York: W.W. Norton, Inc.

Gerber, Alan S., and Donald P. Green. 2008. "Field Experiments and Natural Experiments." *Handbook of Political Methodology* (forthcoming).

Goldthorpe, John. 2001. "Causation, statistics, and sociology." *European Sociological Review* 17 (1): 1-20.

Green, Donald P. 2007. "Experimental Design." *Encyclopedia of Research Methods in the Social Sciences* (forthcoming).

Heckman, James J. 2000. "Causal Parameters and Policy Analysis in Economics: A Twentieth-Century Retrospective." *Quarterly Journal of Economics* 115 (1): 45-97.

Launay, Robert. 2006. “Practical Joking.” In Cécile Canut and Étienne Smith, eds. “Parentés, plaisanteries et politique,” special issue of *Cahiers D'Études Africaines* XLVI (4): 795-808.

Mauss, Marcel. 1928. “Parentés à plaisanterie.” *Annuaire de l'École pratique des hautes études*, section des sciences religieuses (“Les classiques en sciences sociales”) Melun, Imprimerie administrative, Paris: 3-21.

Posner, Daniel N. 2004. “The Political Salience of Cultural Difference: Why Chewas and Tumbukas Are Allies in Zambia and Adversaries in Malawi.” *American Political Science Review* 98 (4): 529-545.

Radcliffe-Brown, A. R. 1940. “On Joking Relationships.” *Africa* 19: 133-140.

Snow, John. 1855. *On the Mode of Communication of Cholera.* Churchill, London. Reprinted in *Snow on Cholera*, London: Humphrey Milford: Oxford University Press, 1936.